Fall 10-31-2016

A Visualization Model Used for Determining the Effectiveness of Information Retrieval in a Scientific Database

Nestor L. Osorio
Northern Illinois University, nosorio@niu.edu

Samantha Johnson
Northern illinois U

A Visualization Model Used for Determining the Effectiveness of Information Retrieval in a Scientific Database

Nestor L. Osorio and **Samantha Johnson**
Northern Illinois University
DeKalb, IL 60115, USA
E-mail: nosorio@niu.edu

Abstract: In this study, we demonstrate that a visualization model can determine the effectiveness of searching a bibliographic database, when three descriptive metadata fields are compared. The Inspec database was searched to create a bibliography of articles about a broad scientific topic, interplanetary travel. We collected metadata from 823 Inspec articles and used the Sci2 Tool to create co-occurrence networks based on subject terms, title keywords, and classification codes from each of the articles. The Watts-Strogatz clustering coefficient model was used to create molecular organization of the networks. This method identified subject domain clusters for each of the three selected metadata elements and subject classification codes were extracted from all the clusters obtained. All data obtained was converted into a common metadata element (classification codes), allowing for the comparison of data from the initial search and from all the subject clusters identified in the visualization process. A set of eight subject codes were found to describe the Main Subject Domain of interplanetary travel. The results also show that searching with classification codes produced the best outcome, the second best option is using subject terms, and the least effective search technique is using title keywords. These results, using visualization, corroborate previous studies.

Keywords: Information retrieval; Knowledge domains; Visualization models; Descriptive metadata; Scientific information.

Introduction

Information retrieval experienced significant growth in the 1980s when online computer-based bibliographic databases became the technology used for searching knowledge domains. The papers of the 1980 symposium on Research and Development in Information Retrieval at the St. John's College in Cambridge, MA, edited by Oddy, Robertson ad Williams (1981) is a compressive summary of the field at that time, with a strong emphasis on definitions, foundations and applications. Several definitions of information retrieval have been presented in the literature, a current one is provided by Ingwersen and Jearvelin, (2005) as: '*The processes involved in representation, storage, searching, finding, filtering and presentation of potential information perceived relevant to a requirement of information desired by a human user in context'*. Further, a significant part of information retrieval research done in the 1980s and 1990s pertains to searching techniques, including the relevance of results; a topic that is central to this project. In this paper, we examine subject keywords in the titles, controlled vocabulary subject terms, and classification codes assigned to citations of a literature search. We use visualization techniques in order to determine subject domains produced by different metadata elements, compare them, and discuss their relevance.

In most cases, visualization methods are used to support computational processes, allowing human perception and cognition to obtain a '*more holistic view of a knowledge domain*' (Skupin, 2014). In this study, we extract data from a visualization process with the purpose of analyzing components of knowledge domains; further, we measure the effectiveness of several searching techniques when used to obtain bibliographic information from a scientific database. In essence, we used human perception and cognition to determine clusters representing components of knowledge domains and subsequently extracted data from the clusters. Therefore, the purpose of this project is three-fold: 1. search a scientific database to create a bibliography of articles about *interplanetary travel*; 2. use a visualization method to determine the subject domains created when searching three different metadata elements obtained from the original search; and 3. extract data from the visualization process to determine the effectiveness of searching a scientific database using the three selected metadata elements.

Information retrieval using title, subject terms, and subject codes

In the 1980s, a great interest in searching techniques was prevalent; a new generation of online database development advanced this topic to a new level. The article on controlled vocabulary by Tenopir (1987) and the review chapter of Travis and Fidel (1982) are classic examples of publications at that time. Tenopir compares the options available for using controlled vocabulary or free-keywords and concludes that '*controlled vocabulary has shown to be the most-effective search method, allowing retrieval of small set of precise items'*. The author also recognized that other factors, such as the subject matter and the characteristics of databases, are determinants for the selection of specific searching methods. Fidel and Travis review the literature of subject analysis covering the period of 1977-1981; free-text searching and controlled vocabulary are the bulk of this review.

Considerable literature on the subject is continuous in the 1990s and 2000s; selected studies are presented in this section. Gomez, Lochbaum and Landauer (1990) study the selection strategies of users performing searches using index-terms. Stokes, Foster and Urguhart (2009) discuss the lack of consensus about methodology for evaluating the retrieval performance of databases; and report that searches are usually performed using titles, keywords, thesaurus terms, and abstracts. Flor, Jakobsson, Mogset, Taylor and Aasen (2001) present the importance of creating a thesaurus for a nursing and allied health database and the process involved in developing a subject thesaurus. Fidel (1991) studies searches done using subject controlled vocabulary terms versus free-subject keywords in a study based on observations of selected searchers; the reasons for selecting thesaurus terms or keyword terms are compiled and analysed.

Papaioannou, Sutton, Carroll, Booth and Wong (2010) concentrate on good practices for searching research papers in the social sciences, they offer searching techniques that include subject searches on multi-databases, but they do not compare subject searches with free-keywords to those done with classification codes. The work of McJunkin (1995), although outdated, is an example of research designed for evaluating the effectiveness of keywords in the title searches. The author uses proximity operators and compares the effectiveness of these operators when searching titles from economics topics in the FirstSearch database. Harmon and Gross (2009) discuss a new taxonomy for titles of scientific documents; they argue that the importance of title formation is critical to identify the '*authors' major claims within a short phase*' and conclude that deleting one significant word from a good title will usually negatively affect the original content. However, the authors make clear than even the best title doesn't reveal processes, methodologies, details of the results, and other important aspects of the research. The work by Voorbij (1998) examines searches performed by experts using title keywords and subject descriptors; their project shows that the relative recall of proper documents was 48% when using title keywords and 86% when using subject descriptors. By conducting a failure analysis, it indicates that although most title keywords were appropriate, they could not retrieve significant documents, a task that was achieved with subject descriptors. Finally, the literature also has a good number of articles using different approaches when searching the online catalogue; the article by McCutcheon (2009) is a good representation of this type of study, the author concludes that using a combination of Library of Congress subject headings and free-keywords can achieve the best results.

Domain analysis using bibliographic data

The study of subject domains is at the centre of *information retrieval*, Shapere (1974), in his work on scientific theories, refers to bodies of knowledge as *domains.* A domain is a body of knowledge clearly defined and is related to other domains in ways such as how electricity is related to magnetism. These bodies of information are characterized by existing association; there are problems (research inquires) on these relationships; the problems (questions) are credible and important; and the problems can have theoretical or practical solutions. Further, Suppe (1989) in his work on the semantic development of theories corroborates Shapere's notion of scientific domains.

Hjørland (2002), who has done some of the pioneering work on domain analysis since the 1990s, published an article that shows how domain analysis can be implemented in the field of information science and proposed eleven approaches. This article is an example of applying domain analysis to a large field of knowledge. With the introduction of new information sources, computational software and visualization techniques, the study of domain analysis has become an important tool for research. Ackerson (1999) establishes relationships between the methods used by scientists for searching the literature of a subject and their perceived mental perception of their field of expertise. Garfield (2004) when studying domains, presents a series of historiographic mappings of knowledge domains which can provide a better understanding of the topic itself. Chen, Paul and O'Keefe (2001) introduced visualization techniques in the study of computer graphics domains based on authors co-citation analysis; and Boyack, Wylie and Davidson (2002) uses VxInsight with relevant bibliographic information that produces examples of domain analyses for the field of science and technology management.

The work of Dutta, Majumder and Sen (2011) is an analytical approach of cluster analysis, by computing parameters such as cluster-rank and cluster-potential from journal articles on brane theory and on fermi liquid. The authors searched those keywords in Inspec and Compendex during the period from 1989 to 2003.The list of citations obtained provides the data collected from elements of the title, abstract, and control vocabulary terms. Further, the article by Stopar, Drobne, Eler and Bartol (2016) explores the field of nanoscience and nanotechnology, by collecting cited and citing articles and using World of Science categories; they determine the different research areas components related to nanoscience and nanotechnology. The authors identify four distinctive groups of categories, as well as measure their interdisciplinarity through citation patterns. The results were obtained by applying scientometric statistical procedures to the data being analysed, which included the presentation of graphical visualization.

Visualization in bibliometric analysis

Data visualization methods are important components for analysing bibliographic information. In *Readings in Information Visualization: Using Vision to Think* by Card, Mackinlay, and Shneiderman (1999) the authors present an exhaustive discussion on visualization, it rationalizes the field, presents classical works produced, analyses applications, and presents its theoretical foundations. Card, et al. define data visualization as '*the use of computer-supported, interactive, visual representations or abstract data to amplify cognition'.* The extensive literature review by Börner, Chen and Boyack (2005) in the *Annual Review of Information Science & Technology* details the initial development of domain visualization. In Meirelles' (2013) discussion of the design of data modelling for information, the author explains that the purpose of visualization can be that of communicating results; documenting research findings and can also be '*a platform for data manipulation and exploration'.* Meirelles also said that the examination of visual displays can be interpreted as a combination of several cognitive elements, among them: the recording of information, the creation of a vehicle for facilitating discovery, the production of models for real or theoretical issues, and the provision for data manipulation.

In this project, we use the text analyser software Sci2 (Science of Science) Tool (2009) capable of producing statistical, temporal, geospatial, topical, and network analyses. We perform database searches which create records of information; by using Sci2 we prepare citations' data in order to execute specific procedures of the software, and we create a real model of subject clusters with the purpose of facilitating discovery. Sci2 is well-recognized software; several research reports using this software have been published, among them: Adams and Light (2014), Belter (2012), Bruer (2010), Cobo, et al. (2011), and Kosecki, Shoemaker and (2011), which show the value of this powerful tool.

Basic concepts of Metadata

As defined by the National Information Standards Organization (2004): '*Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information'.* Three types of metadata have been defined: descriptive, structural, and administrative. Descriptive metadata is the one of concern for this research project; descriptive metadata's purpose is to assign terms to an object (a document in this case) for discovery and identification. Typical descriptive metadata elements are title, subject terms, and subject codes. A database, such as Inspec, is the creator of its own metadata, which is specific for the disciplines that are covered by this database. For a more complete introduction to metadata, the National Information Standards Organization document *Understanding Metadata* is instructive and includes a list of valuable resources. A typical Inspec article citation includes the following descriptive metadata elements: title; authors; source; language; abstract; Inspec headings; key phrase headings; classification; international patent classification; treatment; number of references; publication type; digital object identifier; update code; and accession number. In this paper, the data collected and analysed are from the title, Inspec headings, key phrase headings, and classification (codes).

Data Sources and Processing

Data collection was performed following standard procedures characteristic to bibliometric studies; these practices have been presented, corroborated, and used in numerous studies, De Bellis (2009), Okubo (1997), and Wolfram (2003) among others, in addition to the large number (hundreds) of research-based bibliometric papers published. In this section, we describe the data collection for the first step of the project.

Inspec is one of the most important scientific bibliographic databases. On the EBSCOhost platform, Inspec has an advanced search engine, a thesaurus, and a classification code dictionary. A typical bibliographic citation has the title, authors, authors' affiliations, source, language, and abstract metadata elements, and an additional two metadata elements that provide subject terms: the Inspec headings and the key phrase headings. These two are commonly found in other databases as subject headings or descriptors; each citation also has a list of classification codes. An example of a bibliographic citation is available in appendix 4. Inspec's dictionary of classification codes is an additional tool for indexing and for searching citations; it is divided into five sections: A = physics, B = electrical engineering and electronics, C = computers and control, D = information technology for business, E = production, manufacturing & mechanical engineering. The first three of these sections constitute the major coverage of the database. A classification code is formed by a letter (A, B, C, D, or E) followed by a four digit-number and ends with another letter. The first letter is the section of the database, the first digit is the most general level of classification, the second digit is the second level of classification, and so on. In this paper, we decided to use up to the second level of classification, which provides enough information for the scope of the project. For example, the code B7600 is for aerospace facilities and techniques.

For the purpose of collecting data related to a scientific subject, we selected the topic of *interplanetary travel*, which includes not only complex ideas in areas such as cosmology and astrophysics but also high technological and medical aspects as well. *Interplanetary travel* also represents a very important component of the literature of science fiction. Therefore, we are looking for the scientific and technical aspects of *interplanetary travel*. We searched the Inspec database under the EBSCOhost platform. As indicated previously, Inspec was selected for being one of the most comprehensive bibliographic databases covering all aspects of physics and astronomy; it is equally important for selected areas of engineering and technology. Using the advanced search engine, we used keyword terms *interplanetary travel* or *space travel*. Intentionally, we have used these general keywords terms which cover the topics with the least specificity and allowing for the retrieval of a good number of citations. This first search in Inspec produced 351 citations. We also wanted to explore exotic ideas of space exploration by doing an additional search using the terms *warp or wormholes*. This additional search was motivated by recent articles that discussed the feasibility of these cosmological theories in space travel by Davis (2006 and 2012) where the author presents faster-than-light alternatives for interstellar travel using transversal wormholes or warp drives; these concepts are also important features used in science fiction literature. The search in Inspec for *warp or wormholes* produced 9,289 citations when they were limited to physics related classification terms; the number of citation was 2,115. In order to improve the balance of distribution of citations for the two combined topics, a random sample was taken having a confidence level of 95% and confidence interval of 4% which produced 468 citations Panik (2012). Further, the results of the two previous searches were combined into a single file in the Excel format.

From each citation the following information was initially collected: full bibliographic information, the entire subject field – that in Inspec can be quite large; and the Inspec classification (subject) codes. Subsequently, the data obtained was properly formatted in Excel in four columns: citation number, title, Inspec subject terms, and Inspec classification codes. In order to use the data with Sci2, the Excel file was formatted following the guidelines of Sci2 data preparation and converted into a CSV file. In addition, the title field was edited by selecting only significant subject terms in each title.

The subject field presented a challenge. Inspec uses two types of subject fields: Inspec headings and key phrase headings. Initially, we used all the Inspec headings and key phrase headings attached to each article, which produced an exceedingly large number of subject terms, making the visualization of the clusters difficult to plot and show in table format. A survey of 120 articles, selecting only the Inspec headings, showed that articles with about ten Inspec headings were well described in terms of their subject topics, but those with one or two Inspec headings did not describe their topic well. In order to lessen the number of subject terms, we concluded that collecting the first ten subject terms listed in each citation, whether they were Inspec headings or key phrase headings, would be a good solution to avoid an excessive number of subject terms and properly described each article in the field corresponding to subject terms. The choice made on the number of subject terms also made it more comparable to the number of keywords from the titles available and from the number of assigned Inspec classification subject codes.

The following section presents the process used for obtaining data visualization and the process for extracting data from the clusters obtained.

Methods

Data of subject domains

As mentioned before, Inspec's dictionary of classification codes is divided into five sections: A = physics; B = electrical engineering and electronics; C = computers and control; D = information technology for business; E = production, manufacturing & mechanical engineering. A classification code is formed by a letter (A, B, C, D, or E) followed by a four digit-number and ends with another letter. For example, the code B7600 is for aerospace facilities and techniques. In the next two sections, first we present the procedures used in analysing the data with Sci2, followed by a description of data collection from the visualization model.

Sci2 and method used

Evaluation of data, statistical analysis and the visualization of data was managed by the scientometric analysis and visualization software, Science of Science (Sci2). Sci2 was chosen because of its recognition as one of the most complete tools when it comes to network analysis (Cobo, et al., 2011). Using the Sci2 Tool, three co-occurrence networks were extracted, one from each of the following metadata categories: title keywords, subject terms, and classification codes. Co-occurrence networks work to identify similarities between units (keywords, subject terms, classification codes) by counting the number of times that two units appear together in the specified metadata category across journal articles/entries (Cobo, et al., 2011). Prior to the visualization step, the Watts & Strogatz clustering algorithm was performed on each of the co-occurrence networks to identify modular organization of the networks. '*Small world properties are usually studied to explore networks with tunable values for the average shortest path between pairs of nodes and a high clustering coefficient'*, Watts-Strogatz Small World (2009), p18-19.

The authors also followed the instructions in the Sample Workflows provided by Sci2 Tool (2014). The visualization of the three co-occurrence networks was accomplished using the exploratory data analysis and visualization tool, GUESS. By default, GUESS places nodes in a randomized manner. In order to visualize the appropriate clustering of nodes, the GEM algorithm was applied.

Data analysis

In this section, we present the process of how the data was collected from the visualization procedures used in Sci2. It has two parts: 1. define a knowledge domain from the initial searches on the subject of *interplanetary travel* (we call this domain the *main subject domain*), 2. extract data from the subject clusters identified in the visualization process. The method is descriptive based on the results obtained from the visualization process. We utilize the characteristics of visualization for performing analyses using perception and cognition.

The *main subject domain* was obtained as follows: the two initial and general searches in Inspec using the terms *interplanetary travel* or *space travel* and *warps or wormholes* each generated a set of classification codes titles. They were then converted into classification codes using Inspec's additional tools, such as the thesaurus and classification Index. This conversion procedure was used in order to have a common element of comparison, as will be explained later. The *interplanetary travel* or *space travel* search produced a report of the list of classification code titles that summarize the subject scope of the search. The search for *warps or wormholes* also produced a similar report which was limited to the classification code titles only related to physics or astronomy. *Warps or wormholes* are terms also used in other fields such as textiles, data management, or industrial processes; topics that have no relation to the scientific bibliographic information needed. These two sets of classification code titles were combined into one and constituted the *main subject domain* for *interplanetary travel*. Further, these classification code titles were converted into classification codes.

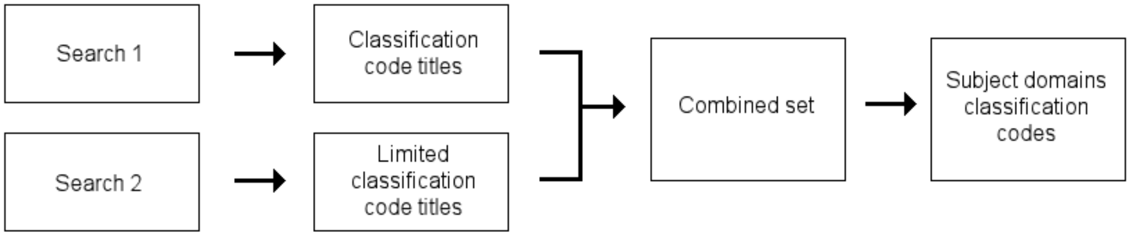

Figure 1: Search 1, data conversion to classification codes

This last set of subject domain classification codes are used as the common element for analysing the results obtained from all the visualization clusters.

As was mentioned before, the visualization process was based on three metadata elements of each of the articles found: keywords from the title, subject terms assigned, and classification codes.

The classification codes from the Sci2 clusters were identified in two ways. For the clusters created in the visualization of classification codes, they were identified using the Layout function of the software and collected in tabular format, see Appendix 2, Tables 11 & 12.

Classification codes from clusters created by the visualization process for 'keywords in the title' and for 'subject terms assigned' were obtained in a similar fashion as those for the *main subject domain* with one additional step: all title keywords and subject terms obtained from the Sci2 clusters were searched for in Inspec, followed by the same steps for finding classification codes titles and classification codes. For example, all title keywords found in cluster one of the Sci2 visualization analysis for title keywords were compiled using the 'OR' function, and a search in Inspec was conducted using the title field. From the results of this search, Inspec generated a list of classification code titles that summarize the subject content of the search. Further, these classification code titles were translated into corresponding classification codes by using the indexing tool in Inspec. A similar process was conducted for clusters created for subject terms.

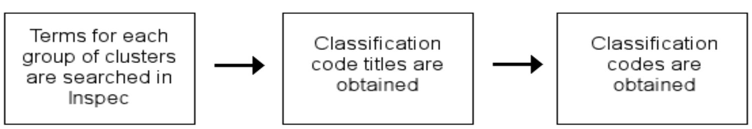

Figure 2: Data collection from clusters

See Appendix 2, tables 7-10. These processes provided us with a single metadata element that is used to compare the effectiveness of searching using title keywords, subject terms, or classification codes.

Results

In the previous section, we have shown the method for converting all data into classification codes, which will be used in this section as a common metadata element of comparison and analysis. It is also recognized that visualization methods produce significant amounts of data; for that reason, the results presented are a descriptive representation of the data collected, limited to the top classification codes identified in each case. By examining all the data produced, we decided that a classification code with three or fewer citations was not included. This limit was adapted only for data collected from the visualization clusters produced by Sci2.

In this section, we present the following results: the set of classification codes and corresponding classification code titles which represent the *main subject domain* of *interplanetary travel*. This is followed by the summary tables corresponding to the clusters created by the visualization procedure for the three metadata elements: keywords from the title, subject terms assigned by Inspec, and classification codes.

Based on the results of the initial search to determine the bibliographic subject components of *interplanetary travel*, the major subject components (according to our search in Inspec) are presented in the following table. We decided to select codes cited up to two times because these codes were obtained directly from the searches performed; a visualization process was not involved, therefore the data obtained was much smaller, see Table 1.

Class Code	Class Code Title
A8700	Biophysics, medical physics, and biomedical engineering (9)
B7600	Aerospace facilities and techniques (7)
A0400	Relativity and gravitation (6)
A9500	Fundamental astronomy and astrophysics, instrumentation & techniques & astronomical observations (6)
B0100	General electrical engineering topics (3)
A1100	General theory of fields and particles (2)
A2800	Nuclear engineering and nuclear power studies (2)
A9800	Stellar systems; Galactic and extragalactic objects and systems; Universe (2)

Table 1: Main Subject Domain of *interplanetary travel*

The table above shows several distinctive areas such as medical physics, aerospace technology, theory of relativity, fields and particles, and cosmology, all related to the current state and future advances of space travel.

Clusters of Subject Domains

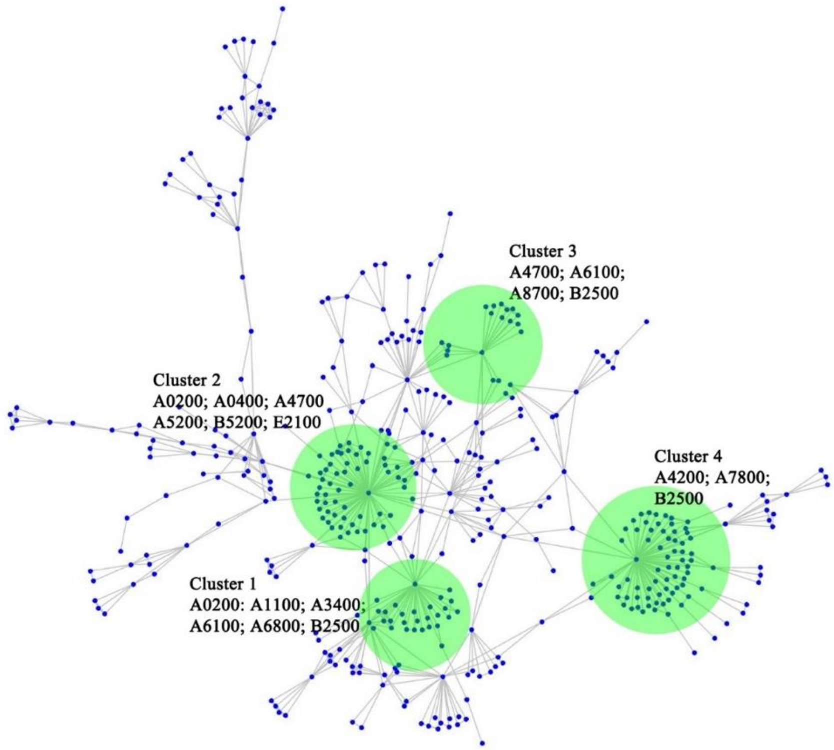

Figure 3: Title keyword clusters

The data collected from the graphs contains a large number of classification codes, in the next following sections, the codes cited three or more times for each subject cluster are presented.

Four clusters were identified when searching with keywords in the title, see Figure 3,

Clusters	Major broader subject codes
Cluster 1	A0200 (4) Mathematical methods in physics; A1100 (4) General theory of fields and particles; A3400 (5) Atomic and molecular collision processes and interactions; A6100 (9) Structure of liquids and solids; crystallography; A6800 (4) Surfaces and interfaces; thin films and whiskers; B2500 (5) Semiconductor materials and devices
Cluster 2	A0200 (5) Mathematical methods in physics; A0400 (4) Relativity and gravitation; A4700 (4) Fluid dynamics; A5200 (6) The physics of plasmas and electric discharges; B5200 (4) Electromagnetic waves, antennas and propagation; E2100 (4) General mechanics
Cluster 3	A4700 (4) Fluid dynamics; A6100 (8) Structure of liquids and solids; crystallography; A8700 (10) Biophysics, medical physics, and biomedical engineering; B2500 (4) Semiconductor materials and devices
Cluster 4	A4200 (14) Optics; A7800 (5) Optical properties and condensed matter spectroscopy and other interactions of matter with particles and radiation; B2500 (10) Semiconductor materials and device

Table 2: keywords in title clusters

Cluster 1 is strong in material sciences and also includes elements of the physics of fields and particles as well as atomic and molecular physics. Cluster 2 includes a variety of scientific and technological topics. Cluster 3 is strong in biophysics and medical physics. Cluster 4 is strong in optical sciences and material science.

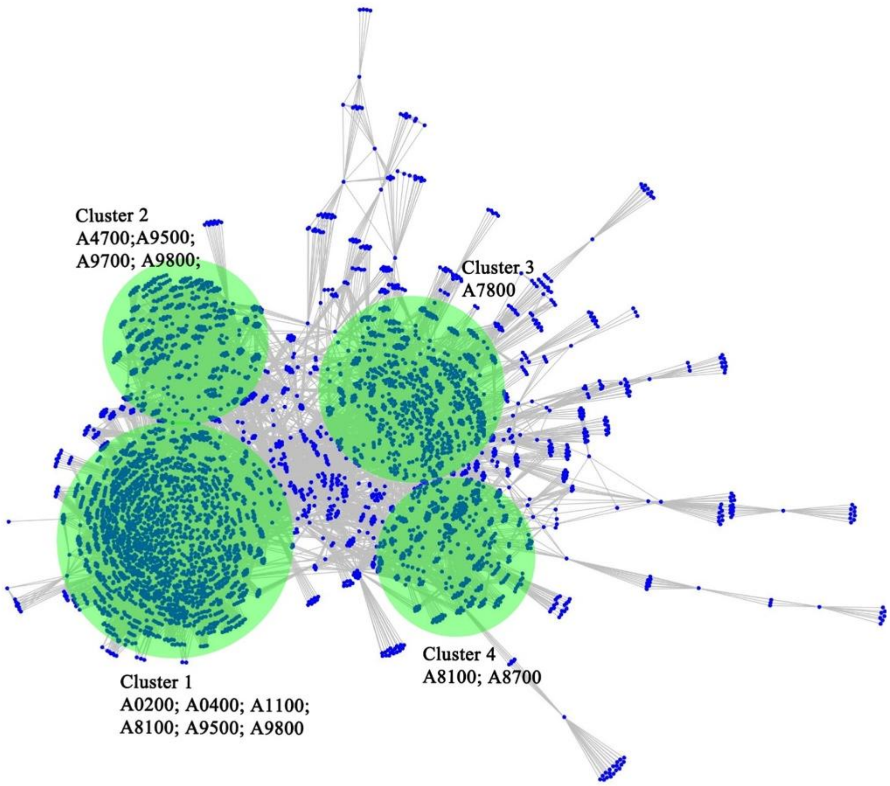

Figure 4: Subject term clusters

Four clusters were identified when searching with subject terms assigned by Inspec, see Figure 4.

Clusters	Major broader subject codes
Cluster 1	A0200 (4) Mathematical methods in physics; A0400 (9) Relativity and gravitation; A1100 (7) General theory of fields and particles; A8100 (3) Materials science; A9500 (3) Fundamental astronomy and astrophysics, instrumentation and techniques and astronomical observations; A9800 (13) Stellar systems; Galactic and extragalactic objects and systems; Universe
Cluster 2	A4700 (4) Fluid dynamics; A9500 (6) Fundamental astronomy and astrophysics, instrumentation and techniques and astronomical observations; A9700 (12) Stars; A9800 (22) Stellar systems; Galactic and extragalactic objects and systems; Universe
Cluster 3	A7800 (5) Optical properties and condensed matter spectroscopy and other interactions; of matter with particles and radiation
Cluster 4	A8100 (4) Materials science; A8700 (23) Biophysics, medical physics, and biomedical engineering

Table 3: Subject terms clusters

Cluster 1 includes the study of relativity, theory of fields, astronomy, astrophysics, and cosmology. Cluster 2 is characterized by the study of stars and cosmology. Cluster 3 is mainly about optical sciences. Cluster 4 has two major components: medical physics and materials science.

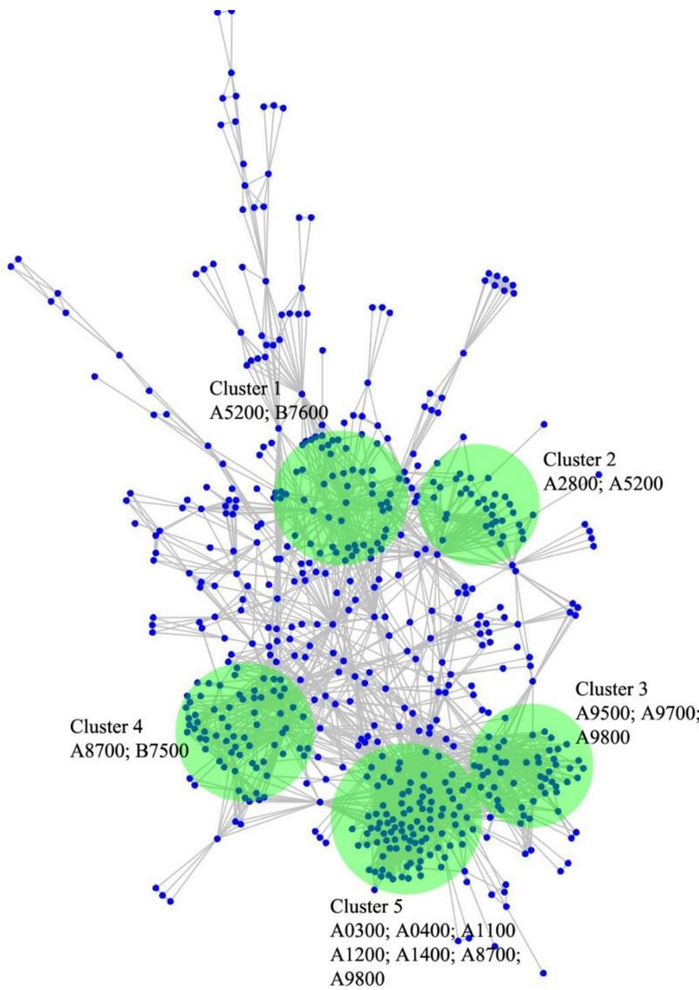

Figure 5: Classification code clusters

Five clusters were identified when searching with Classification Codes, see Figure 5.

Clusters	Major broader subject codes
Cluster 1	A5200 (6) The physics of plasmas and electric discharges; B7600 (4) Aerospace facilities and techniques
Cluster 2	A2800 (7) Nuclear engineering and nuclear power studies; A5200 (5) The physics of plasmas and electric discharges
Cluster 3	A9500 (6) Fundamental astronomy and astrophysics, instrumentation and techniques and astronomical observations; A9700 (13) Stars; A9800 (15) Stellar systems; Galactic and extragalactic objects and systems; Universe
Cluster 4	A8700 (18) Biophysics, medical physics, and biomedical engineering; B7500 (5) Medical physics and biomedical engineering
Cluster 5	A0300 (4) Classical and quantum physics; mechanics and fields; A0400 (5) Statistical physics and thermodynamics; A1100 (7) General theory of fields and particles; A1200 (4) Specific theories and interaction models; particle systematics; A1400 (7) Properties of specific particles and resonances; A8700 (10) Biophysics, medical physics, and biomedical engineering; A9800 (4) Stellar systems; Galactic and extragalactic objects and systems; Universe

Table 4: classification code clusters

The following areas are well-represented in these clusters: cluster 1, physics of plasmas and aerospace technology; cluster 2, nuclear technology and physics of plasmas; cluster 3, the study of stars and cosmology; cluster 4, medical physics and bio-engineering; cluster 5, particle physics and medical physics.

As a whole, the composition of all the three visualization processes presents a well-rounded coverage of the main topic, *interplanetary travel*. In the next section, we discuss how these results compare to the eight main subject components.

Effectiveness of using different metadata elements

In this section, the eight components of the *main subject domain* of *interplanetary travel* are compared to how often those components appear when searching using the three selected metadata elements. For example, A8700 (biophysics, medical physics, and biomedical engineering) appears 10 times in the title keywords search but 28 times in the classification codes search.

Main Components	Title keywords	Subject Terms	Classification Codes
A8700	10	27	28
B7600	0	0	5
A0400	5	9	5
A9500	2	9	2
B0100	0	1	2
A1100	5	7	7
A2800	4	1	8
A9800	4	22	19

Table 5: Comparison of the effectiveness of different searches

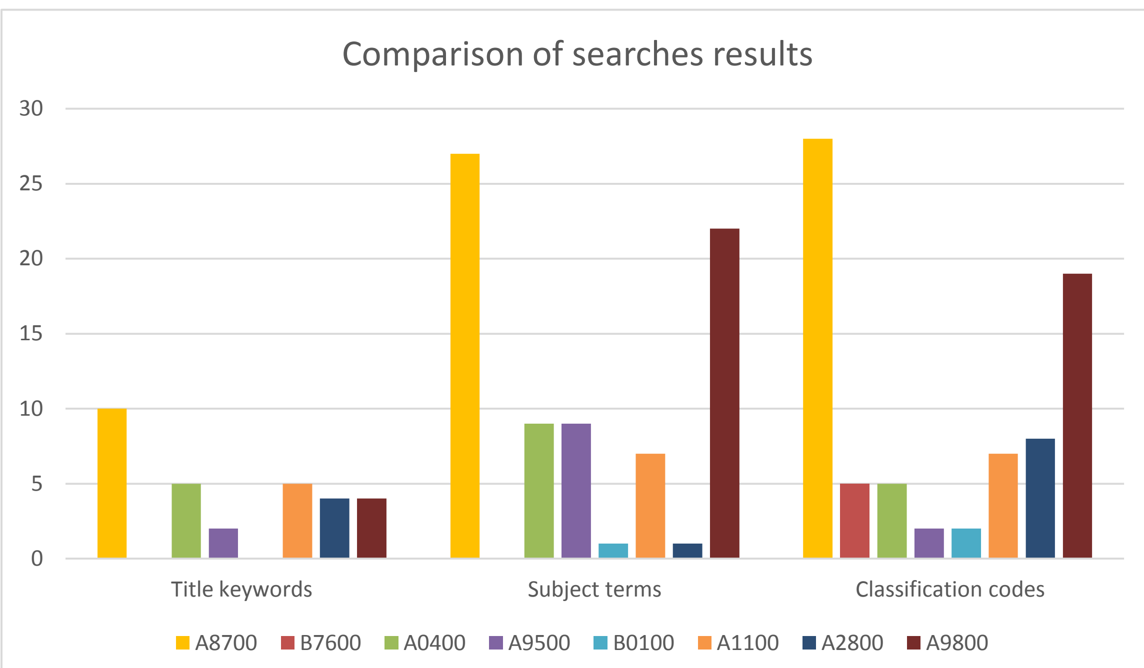

Figure 6: Graphical representation of the effectiveness of searches

Table 5 and Figure 6 summarize the number of times the eight major components of the subject domain for *interplanetary travel* are obtained using the selected metadata elements: keywords in the title, subject terms, and classification codes. These results indicate that searching Inspec using keywords in the title produced the least cited classification codes in number and frequency. Only six of the eight were found, and the frequency of the others is low, except for one (A2800). Searching using subject terms missed only one code (B7600) and the frequencies are in most cases higher than for keywords in the title searches. Searching with classification codes produced cited codes for all areas defined as the *main subject domain* of *interplanetary travel*, as shown in Table 1. This is an indication that searching with classification codes in Inspec is the best choice, although a combination of classification codes and subject terms is a reasonable approach. The least effective search technique is using title keywords. These results corroborate previous database searching studies, such as Voorbij (1998) and Gross, et at. (2015); the results also agree with McCutcheon (2009) which concludes that searching a combination of bibliographic fields (metadata elements) can produce optimal results.

Conclusions

In this study, several methods have been presented with three objectives: 1. search a scientific database to create a bibliography of articles about the scientific topic *interplanetary travel*; 2. use a visualization method to determine the subject domains by searching Inspec using three different metadata elements; and 3. extract data from the visualization process to determine the effectiveness of searching a scientific database using the three selected metadata elements. The main topic *interplanetary travel* was selected only because of its interesting appeal and popularity; any other broader scientific topic could have been used for the same purpose. Objective one was achieved by using the Inspec database, which included both scientific and technological areas in its coverage. Inspec is not totally inclusive and in a real situation, other additional bibliographic databases should be used if a comprehensive coverage of the topic is needed; the Aerospace Database, Compendex, and the Astrophysics Data Systems are three possible selections. Therefore, our results are limited to Inspec.

We have used Sci2, a very powerful visualization software that is well-recognized by researchers, which also has the advantages of being available as an open access package. Therefore, the methods used can be reproduced in any part of the world. We have also introduced an innovative way of using the results of a visualization software by extracting data from the visualization process in order to obtain subject clusters; this data was then used to determine the effectiveness of searching a scientific database using the three selected metadata elements. The results were conclusive; the most inclusive database searches were yielded by searching with classification codes. The use of subject terms was the second best choice, and using title keywords was the least effective choice. The innovative part of this process is that in previous studies, the evaluation of search results was made based on a librarian's or expert's objective evaluation; in this study, we use data from the subject clusters created using Sci2, which allows us to compare the results analytically. The process using Sci2 was labour intensive; we expect that a future version of Sci2 will have the capability to create tabular reports of the components of each cluster; other visualization packages such as T-Lab have that option. Finally, it is understood that not all bibliographic databases have a dictionary of classification codes; further, free-term searching from the abstracts of Inspec's citations was not a technique used in this project. A comparison of free-terms from the abstract and keywords from the title would bring additional insight into searching techniques of bibliographic databases.

About the authors

Nestor L. Osorio is professor and Head of the Reference and Research Department at Northern Illinois University Libraries, DeKalb, Illinois, USA, 60115, nosorio@niu.edu.

Samantha Johnson is a graduate research assistant at Northern Illinois University Libraries and an Audiology Doctoral student at Northern Illinois University, DeKalb, Illinois, USA.

References

Ackerson, L. G. (1999). Visualizing the configuration of scientific literature. *Reference & User Services Quarterly*, 39(1), 43–52.

Adams, J., & Light, R. (2014). Mapping interdisciplinary fields: Efficiencies, gaps and redundancies in HIV/AIDS research. *PLoS ONE*, 9(12), 1–13.

Belter, C, (2012). *Visualizing Networks of Scientific Research*. URL:http://www.infotoday.com/online/may12/Belter-Visualizing-Networks-of-Scientific-Research.shtml. Accessed: 2016-06-05. (Archived by WebCite® at http://www.webcitation.org/6i314GfSf)

Börner, K., Chen, C., & Boyack, K. W. (2005). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179–255.

Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Domain visualization using VxInsight{\textregistered} for science and technology management. *Journal of the American Society for Information Science and Technology*, 53(9), 764–774.

Bruer, J. T. (2010). Can we talk? How the cognitive neuroscience of attention emerged from neurobiology and psychology, 1980-2005. *Scientometrics*, 83(3), 751–764.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings information visualization: Using vision to think*. San Francisco: Morgan Kaufmann.

Chen, C., Paul, R. J., & O'Keefe, B. (2001). Fitting the jigsaw of citation: Information visualization in domain analysis. *Journal of the American Society for Information Science and Technology*, 52(4), 315–330.

Cobo, M. J., Lopez-Herrera, A. G., Herrera-Viedma, E., & Ferrera, F. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402.

Davis, E. (2006). An Assessment of Faster-Than-Light Spacetimes: Make or break issues. *42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, Sacramento, California, AIAA 2006-4908.

Davis, E. (2012). Faster-than-light space warps, status and next steps. *48th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, Joint Propulsion Conferences*, Atlanta, Georgia, AIAA 2006-3860.

De Bellis, N. (2009). *Bibliometrics and citation analysis: from the Science citation index to cybermetrics*. Lanham, Md: Scarecrow Press, 2009.

Dutta, B., Majumder, K., & Sen, B. K. (2011). Study of subject domain by keyword cluster analysis based on research articles: a case study from physics. *Infomation Studies*, 17(4), 195-210.

Fidel, R. (1991). Searchers' Selection of Search Keys: II. Controlled Vocabulary or Free-Text Searching. *Journal of the American Society for Information Science*, 42(7), 501–514.

Flor, P., Jakobsson, A., Mogset, I., Taylor, S., & Aasen, S. E. (2001). A controlled vocabulary for nursing and allied health in Norway. *Health Information and Libraries Journal*, 18(1), 10-19.

Garfield, E. (2004). Historiographic mapping of knowledge domains literature, *Journal of Information Science*, 30(2), 119-145.

Gross, T., Taylor, A. G. & Joudrey, D. (2015). Still a lot to lose: The role of controlled vocabulary in keyword searching. *Cataloging & Classification Quarterly*, 53(1), 1-39.

Gomez, L. M., Lochbaum, C. C., & Landauer, T. K. (1990). All the right words: Finding what you want as a function of richness of indexing vocabulary. *Journal of the American Society for Information Science*, 41(8), 547–559.

Harmon, J. E., & Gross, A. G. (2009). The structure of scientific titles. *Journal of Technical Writing and Communication*, 39(4), 455–465.

Hjørland, B. (2002). Domain analysis in information science. *Journal of Documentation*, 58(4), 422–462.

Ingwersen, P., & Jearvelin, K. (2005). *The turn: integration of information seeking and retrieval in context*. Dordrecht: Springer.

Kosecki, S., Shoemaker, R., & Baer, C. K. (2011). Scope, characteristics, and use of the U.S. Department of Agriculture's intramural research. *Scientometrics*, 88(3), 707–728.

McCutcheon, S. (2009). Keyword vs controlled vocabulary searching: the one with the most tools wins. *Indexer*, 27(2), 62–65.

McJunkin, M. C. (1995). Precision and recall in title keyword searches. *Information Technology and Libraries*, 14(3), 161–171.

Meirelles, I. (2013). *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations*. Beverly, MA: Rockport Publishers.

National Infomation Standards Organization. (2004). *Understanding metadata*. NISO Press. URL:http://www.niso.org/publications/press/UnderstandingMetadata.pdf . Accessed: 2016-06-05. (Archived by WebCite® at http://www.webcitation.org/6i34mQOAJ)

Oddy, R. N., Robertson, S. E., & Williams, P. W. (1981). *Information retrieval research*. London; Butterworths.

Okubo, Y. (1997). Bibliometric indicators and analysis of research systems: Methods and examples. OECD Science, *Technology and Industry Working Papers*, 1997(01), 1–70.

Panik, M. J. (2012). *Statistical inference: A short course*. Hoboken, NJ: Wiley

Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010). Literature searching for social science systematic reviews: Consideration of a range of search techniques. *Health Information and Libraries Journal*, 27(2), 114–122.

Sci2 Team. (2009). *Science of Science (Sci2) Tool*. Indiana University and SciTech Strategies. URL:http://sci2.cns.iu.edu. Accessed: 2016-06-05. (Archived by WebCite® at http://www.webcitation.org/6i36EovPd)

Sci2 Team. (2014). *Science of Science (Sci2) Tool Manual*, v1.1 beta. Blomington, IN: SciTech Strategies Inc. URL:http://wiki.cns.iu.edu/display/SCI2TUTORIAL/Science+of+Science+%28Sci2%29+Tool+Manual. Accessed: 2016-06-05. (Archived by WebCite® at http://www.webcitation.org/6i38beurg)

Shapere, D. (1974). Scientific theories and their domains. In F. Suppe, *The structure of scientific theories*. Urbana: University of Illinois Press.

Skupin, A. (2014). Making a mark: a computational and visual analysis of one researcher's intellectual domain. *International Journal of Geographical Information Science*, 28(6), 1209–1232.

Stokes, P., Foster, A., & Urquhart, C. (2009). Beyond relevance and recall: Testing new user-centred measures of database performance. *Health Information and Libraries Journal*, 26(3), 220–231.

Stopar, K., Drobne, D., Eler, K., & Bartol, T. (2016). Citation analysis and mapping of nanoscience and nanotechnology: identifying the scope and interdisciplinarity of research. *Scientometrics*, 106(2), 563–581.

Suppe, F. (1989). *The semantic conception of theories and scientific realism*. Urbana: University of Illinois Press.

Tenopir, C. (1987). Searching by controlled vocabulary or free text? *Library Journal*, 112(19), 58–59.

Travis, I. L., & Fidel, R. (1982). 5 Subject analysis. *Annual Review of Information Science and Technology*, 17, 123–157.

Voorbij, H. J. (1998). Title keywords and subject descriptors: a comparison of subject search entries of books in the humanities and social sciences. *Journal of Documentation*, 54(4), 466–476.

Watts-Strogatz Small World. (2009). In *Network workbench tool. User manual 1.0.0,* p. 18-19. Indiana University, Bloomington, IN: Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University. URL:http://nwb.cns.iu.edu/Docs/NWBTool-Manual.pdf. Accessed: 2016-06-05. (Archived by WebCite® at http://www.webcitation.org/6i39QiSE7)

Wolfram, D. (2003). *Applied informetrics for information retrieval research*. Westport, CT: Libraries Unlimited.

Appendixes

Appendix 1

interplanetary travel or 'space travel' – classification codes	warps or wormholes - classification codes	summary of all codes	all codes
A8700	A9800	A0100	A0130R
B7600	A9500	A0300	A0330
A8700	A0400	A0400	A0400
B7600	A0400		A0400
E3600	A0400		A0400
A9500	A9800		A0400
A9500	A1100		A0400
A8700	A0400		A0420
B7600	A9700	A1100	A1100
C3300	A1100		A1100
B 7600	A0400	A2800	A2850P
A9400	A9500		A2852J
B7600		A5200	A5275D
A8700		A8100	A8180
A8700		A8600	A8630J
C7400		A8700	A8700
E2320			A8700
A8700			A8700
B7600			A8700
A8700			A8700
E1020			A8700
A8180			A8700
A0130R			A8770E
E3050			A8790
A5275D		A9400	A9400
A0420		A9500	A9500
A8790			A9500
A8770E			A9500
B7530B			A9500
B0170N			A9530S
C7810C			A9555W
E1710		A9700	A9700
B0170G		A9800	A9800
B76905			A9800
A8630J		B0100	B0160
A9530S			B0170G
B6250G			B0170N
B8420		B6200	B6250G
A0330		B7500	B7530B
E1630		B7600	B7600
B8520			B7600
A9555W			B7600
C7350			B7600
C7330			B7600
A2852J			B7600
E0240H			B76905
A2850P		B8200	B8220B
B8220B		B8400	B8420
B0160		B8500	B8520
C6180		C3300	C3300
		C6100	C6180
		C7300	C7330
		C7300	C7350
		C7400	C7400
		C7800	C7810C

		E0200	E0240H
		E1000	E1020
		E1600	E1630
		E1700	E1710
		E2300	E2320
		E3000	E3050
		E3600	E3600

Table 6. Classification Codes of the Main Subject Domain for *interplanetary travel*

Appendix 2

This appendix includes all the data obtained from the visualization process.

Keywords in the Title Clusters

Elements of each of the four clusters are presented. The first column lists broader codes, the second column lists the codes obtained from the searching procedure explained earlier and represents the final conversion to classification codes.

Title keywords Cluster 1		Title keywords Cluster 2		Title keywords Cluster 3		Title keywords Cluster 4	
A0200	A0210	A0200	A0210	A0200	A0230	A4200	A4225F
	A0230		A0230		A0260		A4250
	A0240		A0240	A2500	A2550G		A4255P
	A0260		A0250	A2700	A2710		A4260B
A0300	A0340K		A0260	A2800	A2852C		A4260D
	A0365G	A0300	A0340D		A2852F		A4260F
A0400	A0420J		A0365B		A2852J		A4260H
A0700	A0777	A0400	A0420J	A4400	A4440		A4265
	A0785		A0455	A4700	A4710		A4265
A1100	A1110N		A0460		A4725		A4265K
	A1117		A0470		A4755M		A4272
	A1130P	A1100	A1110N		A4760		A4280K
	A1130R	A2800	A2852J	A5200	A5255G		A4280L
A1200	A1210	A4100	A4110H		A5255P		A4280W
A2900	A2925F	A4200	A4250	A6100	A6140K	A6100	A6146
A3100	A3150	A4300	A4320		A6170B	A6800	A6855
A3400	A3450		A4340		A6180		A6865
	A3450H	A4600	A4630J		A6180B	A7100	A7135
	A3480B		A4630M		A6180E	A7200	A7220J
	A3480D	A4700	A4710		A6180F		A7240
	A3480G		A4725		A6180H	A7300	A7320D
A4100	A4110H		A4730		A6180J		A7340L
A6100	A6114H		A4760	A6200	A6220F	A7800	A7820D
	A6146	A5200	A5225F		A6220M		A7830G
	A6170B		A5225L	A6400	A6475		A7835
	A6170T		A5230	A6800	A6855		A7840H
	A6170T		A5235P	A7800	A7820D		A7865K
	A6170W		A5255G	A8100	A8140G	A8200	A8265J
	A6180F		A5265		A8140L	A8600	A8630J
	A6180J	A6200	A6220D		A8140N	B2500	B2520C
	A6180M	A7300	A7320D	A8200	A8250		B2520D
A6800	A6820		A7340G	A8700	A8715M		B2530B
	A6845B	A8100	A8140L		A8716		B2530C
	A6845D		A8140L		A8725		B2530F
	A6855	A9500	A9530S		A8725F		B2550E
A7900	A7920K	A9800	A9880B		A8750		B2560B
	A7920N		A9880D		A8750E		B2560P
A8100	A8160C		A9880L		A8750G		B2560R

	A8230F	B0200	B0220		A8760J		B2570
A9500	A9530S		B0290		A8760M	B4100	B4110
A9800	A9880D	B0500	B0550		A8770H		B4125
B2500	B2520C	B1300	B1320	A9200	A9260W		B4130
	B2520D	B5200	B5210	B0500	B0550	B4200	B4260D
	B2550A		B5240D	B2500	B2520D	B4300	B4320J
	B2550A		B5260		B2550R	B6100	B6135
	B2550E		B5270B		B2560R	B6200	B6260C
B5200	B5210	E2100	E2110A		B2560S	B7200	B7230C
B7400	B7410B		E2180A	B5200	B5270B	B8400	B8420
	B7410D		E2180B	B7400	B7420	C5200	C5260B
	B7420		E2180C	B7500	B7530B	E3600	E3644A

Table 7. Keywords in the Title Clusters

In column 3, one item has two codes: A8770H - B7520C; the first code was selected. In column 4, one item has two codes: A4260F/ B4330B; the first code was selected.

Cluster	Broader codes
Cluster 1	A0200 (4), A0300 (2), A0400 (1), A0700 (2), A1110 (4), A1200 (1), A2900 (1), A3100 (1), A3400 (5), A4100 (1), A6100 (9), A6800 (4), A7900 (2), A8100 (2), A9500 (1), A9800 (1), B2500 (5), B5200 (1), B7400 (3)
Cluster 2	A0200 (5), A0300 (2), A0400 (4), A1100 (1), A2800 (1), A4100 (1), A4200 (1), A4300 (2), A4600 (2), A4700 (4), A5200 (6), A6200 (1), A7300 (2), A8100 (2), A9500 (1), A9800 (3), B0200 (2), B0500 (1), B1300 (1), B5200 (4), E2100 (4)
Cluster 3	A0200 (2), A2500 (1), A2700 (1), A2800 (3), A4400 (1), A4700 (4), A5200 (2), A6100 (8), A6200 (2), A6400 (1), A6800 (1), A7800 (1), A8100 (3), A8200 (1), A8700 (10), A9200 (1), B0500 (1), B2500 (4), B5200 (1), B7400 (1), B7500 (1)
Cluster 4	A4200 (14), A6100 (1), A6800 (2), A7100 (1), A7200 (2), A7300 (2), A7800 (5), A8200 (1), A8600 (1), B2500 (10), B4100 (3), B4200 (1), B4300 (1), B6100 (1), B6200 (1), B7200 (1), B8400 (1), C5200 (1), E3600 (1)

Table 8. Keywords in the title clusters with frequency for each element

Subject Terms Clusters

Elements of each of the four clusters are presented. The first column lists broader codes, the second column lists the codes obtained from the searching procedure explained earlier and represents the final conversion to classification codes.

Subject terms Cluster 1		Subject terms Cluster 2		Subject terms Cluster 3		Subject terms Cluster 4	
A0200	A0210	A0200	A0260	A0200	A0260	A0200	A0250
	A0230	A4700	A4710	A2800	A2844		A0260
	A0240		A4755K	A4200	A4255P	A6100	A6180J
	A0260		A4760		A4260B	A6200	A6220M
A0400	A0420C		A4785	A4700	A4710	A6400	A6480G
	A0420J	A8700	A8760F		A4760	A6800	A6855
	A0420J		A8770E	A6100	A6146	A8100	A8140G
	A0430	A9500	A9530L	A6400	A6480G		A8140L
	A0440		A9530S	A6800	A6820		A8140L
	A0455		A9580D		A6855		A8140N
	A0460		A9580G		A6865	A8200	A8245
	A0465		A9580J	A7400	A7470V	A8700	A8710
	A0470		A9580N	A7800	A7820D		A8715
A0500	A0570C	A9700	A9710B		A7830G		A8715B
A1100	A1110C		A9710C		A7840H		A8715D
	A1110N		A9710E		A7855H		A8715D
	A1117		A9710F		A7865K		A8715H

	A1130E		A9710H	A8100	A8116		A8715K
	A1130J		A9710N		A8140G		A8715M
	A1130P		A9710Q	A8200	A8245		A8725
	A1130Q		A9710R	A8700	A8760F		A8725F
A1200	A1210		A9710T		A8770E		A8730
	A1210		A9710W	A9300	A9385		A8730C
	A1210D		A9720L	B0100	B0160		A8745B
A6100	A6146		A9760B	B0200	B0240Z		A8745D
A6200	A6220F	A9800	A9810		B0290		A8745H
A6400	A6470K		A9820E	B0500	B0520F		A8750G
	A6480G		A9820G		B0550		A8760F
A8100	A8130H		A9840B	B2500	B2520D		A8760I
	A8140G		A9840C	B4300	B4320J		A8760J
	A8140L		A9840K	B6100	B6135		A8770
A9500	A9530C		A9840L	B7200	B7230		A8770E
	A9530L		A9850B	B8400	B8410E		A8770M
	A9530S		A9850D	B8500	B8520		A8780
A9700	A9710R		A9850E	C1100	C1160	B0200	B0240Z
	A9760L		A9850F		C1180	B0500	B0550
	A9780J		A9850G	C1200	C1290H	B6100	B6135
A9800	A9850B		A9850H		C1290P		B6135E
	A9850E		A9850K	C3100	C3120C	B7400	B7410B
	A9850G		A9850L	C3300	C3360L	B7500	B7510H
	A9850H		A9850M		C3390		B7510J
	A9850K		A9850R		C3390C		B7520
	A9850R		A9850T	C5200	C5260B	C1100	C1160
	A9870J		A9870D	C7300	C7330		C1180
	A9870Q		A9870Q	C7400	C7420	C5200	C5260B
	A9870V		A9880D	C7800	C7840	C6100	C6170K
	A9880B		A9880L	E0200	E0230		C6180
	A9880D	B0200	B0290		E0240H	C7300	C7330
	A9880D	B7500	B7510J	E1000	E1020	E1400	E1410
	A9880L	E2100	E2130	E3600	E3650C	E1500	E1525

Table 9. Subject Terms Clusters

Cluster	Broader codes
Cluster 1	A0200 (4), A0400 (9), A0500 (1), A1100 (7), A1200 (3), A6100 (1), A6200 (1), A6400 (2), A8100 (3), A9500 (3), A9700 (3), A9800 (13)
Cluster 2	A0200 (1), A4700 (4), A8700 (2), A9500 (6), A9700 (12), A9800 (22), B0200 (1), B7500 (1), E2100 (1)
Cluster 3	A0200 (1), A2800 (1), A4200 (2), A4700 (2), A6100 (1), A6400 (1), A6800 (3), A7400 (1), A7800 (5), A8100 (2), A8200 (1), A8700 (2), A9300 (1), B0100 (1), B0200 (2), B0500 (2), B2500 (1), B4300 (1), B6100 (1), B7200 (1), B8400 (1), B8500 (1), C1100 (2), C1200 (2), C3100 (1), C3300 (3), C5200 (1), C7300 (1), C7400 (1), C7800 (1), E0200 (2), E1000 (1), E3600 (1)
Cluster 4	A0200 (2), A6100 (1), A6200 (1), A6400 (1), A6800 (1), A8100 (4), A8200 (), A8700 (23), B0200 (1), B0500 (1), B6100 (2), B7400 (1), B7500 (3), C1100 (2), C5200 (1), C6100 (2), C7300 (1), E1400 (1), E1500 (1)

Table 10. Subject terms clusters with frequency for each element

Classification Codes clusters

Elements of each of the five clusters are presented. The first column shows the broader codes, the second column lists the codes found in the Sci2 visualization graphs.

Custer 1		Cluster 2		Custer 3		Cluster 4		Custer 5	
A4100 A5200	A4160g A5200	A0700	A0730 A0750	A0100 A9400	A0140d A9440	A7500 A8100	A7580 A8120	A0100	A0130g A0130r

	A5235p	A2800	A2841	A9500	A9510	A8200	A8235		A0138l
	A5240k		A2841d		A9510j	A8700	A8710	A0200	A0220
	A5255m		A2850k		A9580d		A8715g		A0240
	A5265		A2852		A9580g		A8715k	A0300	A0360
	A5275		A2852c		A9580j		A8730g		A0365B
	A5280p		A2852l		A9580n		A8732		A0365g
A8200	A8240p		A2868e	A9600	A9610		A8745		A0367
A8600	A8610n	A4700	A4725f	A9700	A9710		A8745f	A0400	A0420
	A8630	A5200	A5250		A9710		A8745h		A0420j
	A8630m		A5250g		A9710k		A8745h		A0420m
A9500	A9575		A5250j		A9710r		A8750b		A0430
A9600	A9660		A5255k		A9720e		A8760		A0440
B0100	B0140B		A5270n		A9720j		A8760b	A1100	A1100g
	B0170n	A7900	A7920d		A9720l		A8760f		A1110
B7600	B7630	A9300	A9385		A9720t		A8760j		A1110j
	B7630	B2500	B2560z		A9730e		A8770		A1110l
	B7630B	B4300	B4360B		A9730q		A8770e		A1130C
	B7640	B6100	B6120		A9780f		A8770h		A1130g
B7700	B7710	B7600	B7600		A9780g		A8780		A1130j
B7900	B7910	B7900	B7990		A9780j	A9200	A9220j	A1200	A1210B
B8400	B8420	B8200	B8220	A9800	A9810	B0200	B0240g		A1210d
	B8460	B8300	B8360		A9820e	B0500	B0580		A1220d
B8500	B8520				A9840	B6100	B6135		A1220d
C1100	C1120				A9840b	B6200	B6210	A1300	A1330
	C1180				A9840c	B7200	B7210x		A1330g
C1300	C1320				A9840f	B7500	B7510	A1400	A1440d
	C1340d				A9840k		B7510n		A1460g
C3300	C3360l				A9850d		B7520c		A1480A
C4100	C4190				A9850e		B7550		A1480d
E0100	E0120k				A9850f		B7580		A1480j
E0200	E0240h				A9850k	C5200	C5260b		A1480k
	E0260				A9850l	C6100	C6130m		A1480l
E1000	E1020				A9850r		C6170t	A2800	A2880f
E2100	E2120				A9870d	C7800	C7820	A6400	A6485
E2300	E2320				A9870q		C7830	A8700	A8725
E3600	E3650C			B7400	Bb7420c				A8730
									A8730C
									A8732e
									A8732s
									A8760m
									A8760m
									A8760p
									A8765
									A8790
								A9400	A9440e
									A9440r
								A9500	A9530g
									A9575p
								A9700	A9710q
									A9760j
									A9760s
								A9800	A9800B
									A9850h
									A9880
									A9880
								B7200	B7230
								B7500	B7530
								C4200	C4260
								C6100	C6130B
								C7300	C7350
								C7800	C7810
									C7810

Table 11. Classification Codes Clusters

Cluster	Broader codes
Cluster 1	A4100 (1), A5200 (7), A8200 (1), A8300 (1), A8600 (3), A9500 (1), A9600 (1), B0100 (2), B7600 (4), B7700 (1), B7900 (1), B8400 (2), B8500 (1), C1100 (2), C1300 (2), C3300 (1), C4100 (1), E0100 (1), E0200 (2), E1000 (1), E2100 (1), E2300 (1), E3600 (1)
Cluster 2	A0700 (2), A2800 (7), A4700 (1), A5200 (5), A7900 (1), A9300 (1), B2500 (1), B4300 (1), B6100 (1), B7600 (1), B7900 (1), B8200 (1), B8300 (1)
Cluster 3	A0100 (1), A9400 (1), A9500 (6), A9600 (1), A9700 (13), A9800 (15), B7400 (1)
Cluster 4	A7500 (1), A8100 (1), A8200 (1), A8700 (18), A9200 (1), B0200 (1), B0500 (1), B6100 (1), B6200 (1), B7200 (1), B7500 (5), C5200 (1), C6100 (2), C7800 (1)
Cluster 5	A0100 (3), A0200 (2), A0300 (4), A0400 (5), A1100 (7), A1200 (4), A1300 (2), A1400 (7), A2800 (1), A6400 (1), A8700 (10), A9400 (2), A9500 (2), A9700 (3), A9800 (4), B7200 (1), B7500 (1),C4200 (1), C6100 (1), C7300 (1), C7800 (2)

Table 12. Classification codes clusters with frequency for each element

Numbers in parenthesis show the number of times a code was identified.

Appendix 3

All the major classification codes identified in this article are listed in the following table.

Classification code	Classification code title
A0100	Communication, education, history, and philosophy
A0200	Mathematical methods in physics
A0300	Classical and quantum physics; mechanics and fields
A0400	Relativity and gravitation
A0500	Statistical physics and thermodynamics
A0700	Specific instrumentation and techniques of general use in physics
A1100	General theory of fields and particles
A1200	Specific theories and interaction models; particle systematics
A1300	Specific elementary particle reactions and phenomenology
A1400	Properties of specific particles and resonances
A2500	Nuclear reactions and scattering: specific reactions
A2700	Properties of specific nuclei listed by mass ranges
A2800	Nuclear engineering and nuclear power studies
A2900	Experimental methods and instrumentation for elementary-particle and nuclear physics
A3100	Theory of atoms and molecules
A3400	Atomic and molecular collision processes and interactions
A4100	Electricity and magnetism; fields and charged particles
A4200	Optics
A4300	Acoustics
A4400	Heat flow, thermal and thermodynamic processes
A4600	Mechanics, elasticity, rheology
A4700	Fluid dynamics
A5200	The physics of plasmas and electric discharges
A6100	Structure of liquids and solids; crystallography
A6200	Mechanical and acoustic properties of condensed matter
A6400	Equations of state, phase equilibria, and phase transitions
A6800	Surfaces and interfaces; thin films and whiskers
A7100	Electron states in condensed matter
A7200	Electronic transport in condensed matter
A7300	Electronic structure and electrical properties of surfaces, interfaces, and thin films
A7400	Superconductivity

A7500	Magnetic properties and materials
A7800	Optical properties and condensed matter spectroscopy and other interactions of matter with particles and radiation
A7900	Electron and ion emission by liquids and solids; impact phenomena
A8100	Materials science
A8200	Physical chemistry
A8600	Energy research and environmental science
A8700	Biophysics, medical physics, and biomedical engineering
A9200	Hydrospheric and lower atmospheric physics
A9300	Geophysical observations, instrumentation, and techniques
A9500	Fundamental astronomy and astrophysics, instrumentation and techniques and astronomical observations
A9600	Solar system
A9700	Stars
A9800	Stellar systems; Galactic and extragalactic objects and systems; Universe
B0100	General electrical engineering topics
B0200	Engineering mathematics and mathematical techniques
B0500	Materials science for electrical and electronic engineering
B1300	Microwave technology
B2500	Semiconductor materials and devices
B4100	Optical materials and devices
B4200	Optoelectronic materials and devices
B4300	Lasers and masers
B5200	Electromagnetic waves, antennas and propagation
B6100	Information and communication theory
B6200	Telecommunication
B7200	Measurement equipment and instrumentation systems
B7400	Elementary particle and nuclear instrumentation
B7500	Medical physics and biomedical engineering
B7600	Aerospace facilities and techniques
B7700	Earth sciences
B7900	Military systems and equipment
B8200	Generating stations and plants
B8300	Power apparatus and electric machines
B8400	Direct energy conversion and energy storage
B8500	Power utilisation
C1100	Mathematical techniques
C1200	Systems theory and cybernetics
C1300	Control theory
C3100	Control and measurement of specific variables
C3300	Control applications
C4100	Numerical analysis
C4200	Computer theory
C5200	Logic design and digital techniques
C6100	Software techniques and systems
C7300	Natural sciences computing
C7400	Engineering computing
C7800	Other computer applications
E0100	Management and business
E0200	General support functions
E1000	Manufacturing and production
E1400	Design
E1500	Production technology
E2100	General mechanics
E2300	Mechanical machinery
E3600	Manufacturing industries

Table 13. Inspec Classification Codes with respective Classification Codes Titles

Appendix 4

Inspec bibliographic citation

Wormholes, time machines, and the weak energy condition

Authors: Morris, M.S.[1]; Thorne, K.S.[1]; Yurtsever, U.[1]

Author's Affiliation: [1]Theor. Astrophys., California Inst. of Technol., Pasadena, CA, USA

Source: Physical Review Letters 26 Sept. 1988, vol.61, no.13, pp. 1446-9. *ISSN:* 0031-9007 (print), *CODEN:* PRLTAO *Country of Publication:* USA

Language: English

Abstract: It is argued that, if the laws of physics permit an advanced civilization to create and maintain a **wormhole** in **space** for interstellar **travel**, then that **wormhole** can be converted into a time machine with which causality might be violatable. Whether **wormholes** can be created and maintained entails deep, ill-understood issues about cosmic censorship, quantum gravity, and quantum field theory, including the question of whether field theory enforces an averaged version of the weak energy condition.

Inspec Headings: cosmology; general relativity; quantum field theory of gravitation; Schwarzschild metric

Key Phrase Headings: general relativity; Schwarzschild metric; **wormhole**; interstellar **travel**; time machine; causality; cosmic censorship; quantum gravity; quantum field theory; weak energy condition

Classification: A0420C Fundamental problems and general formalism in general relativity
A0460 Quantum theory of gravitation
A9530S Relativity and gravitation in astrophysics
A9880 Cosmology

Treatment: Theoretical or Mathematical

Number of References: 18

Publication Type: Journal Paper

Update Code: 1989002

Accession Number: 3274490

Figure 7: Inspec bibliographic record

Appendix 5

Graphs of the cluster, showing labels:

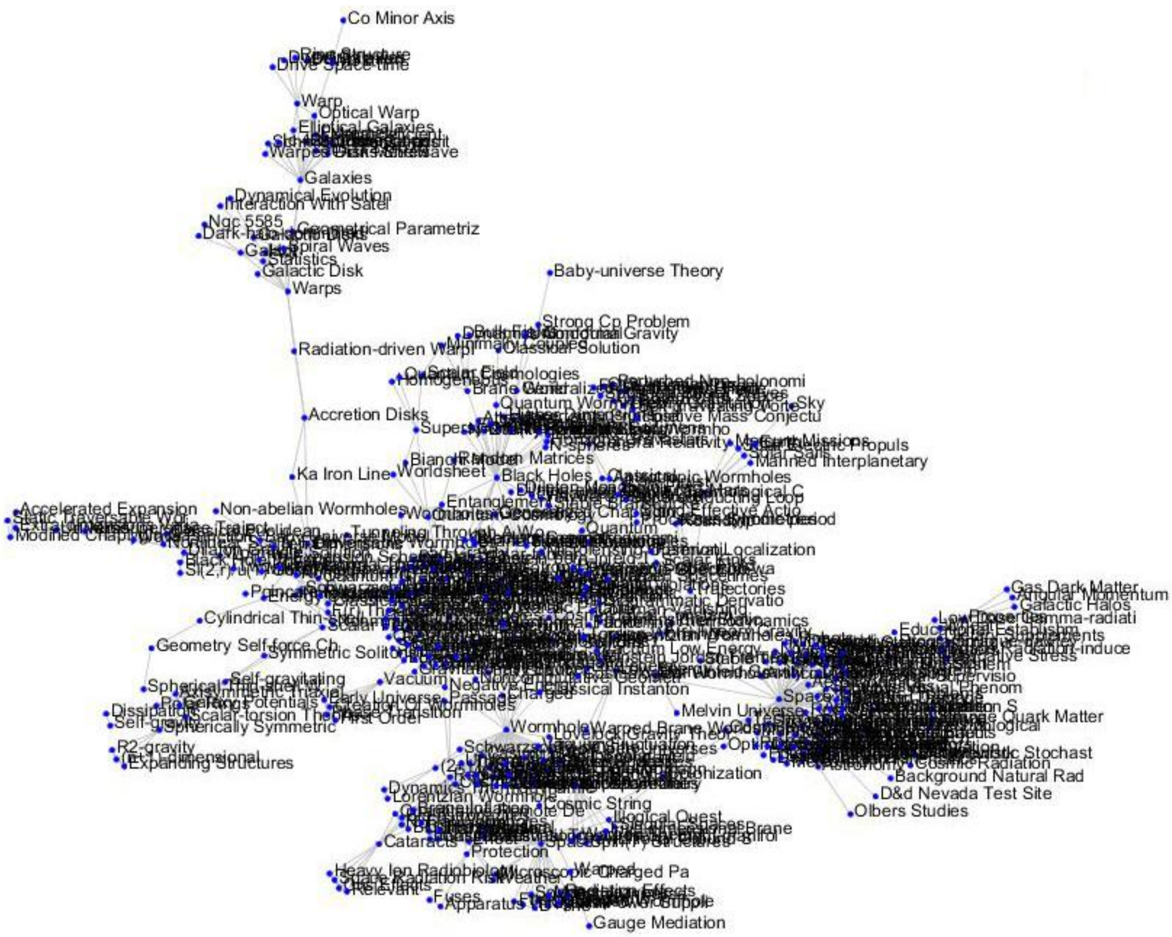

Figure 8: Title keyword clusters

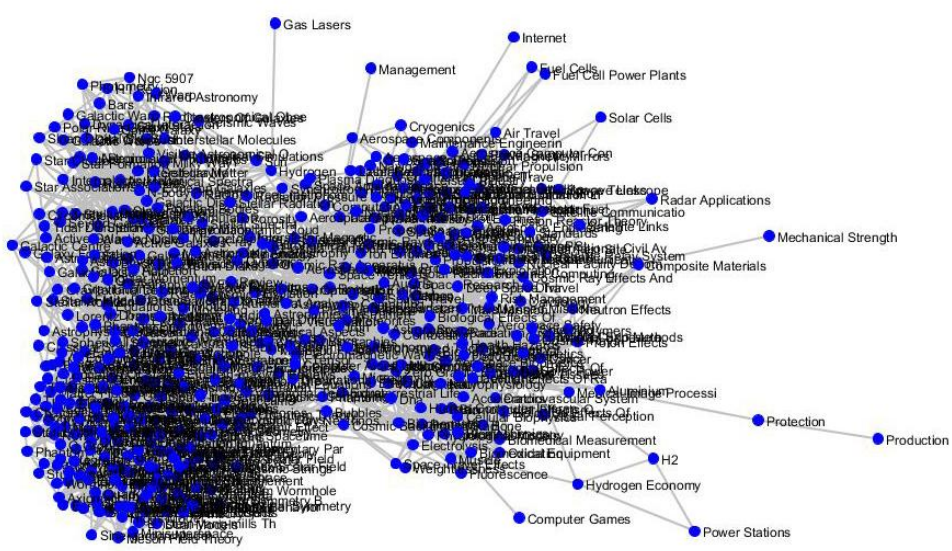

Figure 9: Subject terms clusters*

*This visualization has been limited to 500 subject term nodes for the purpose of clarity.

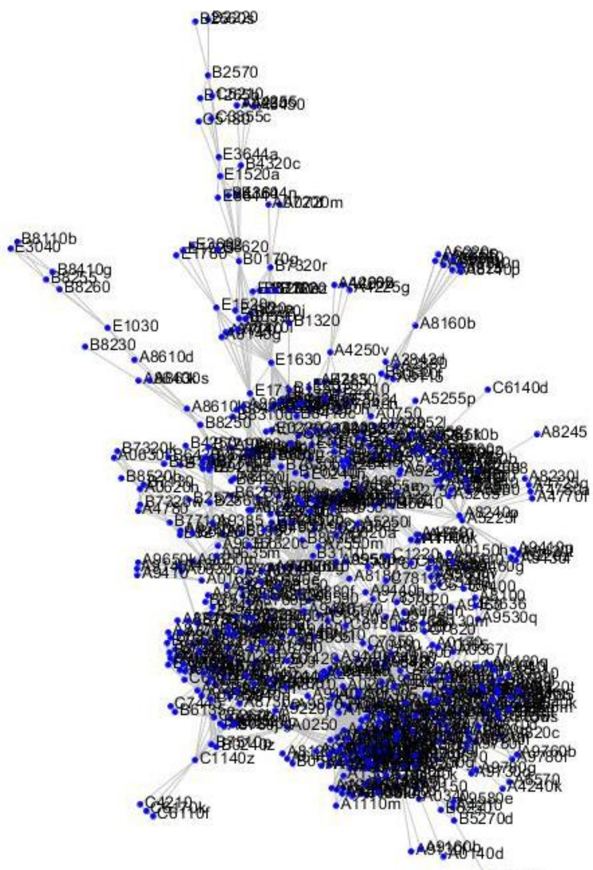

Figure 10: Classification Codes clusters